Spike and Slab Priors in Practice

A practical guide to high dimensional inference

Michael Komodromos

November 29, 2021

Outline

- Recap
  - Problem
  - Spike and Slab priors
- Practical concerns
  - Model selection
  - Tuning hyperparameters
- An example on simulated data
  - Regularization paths
- A case study on real data
  - TCGA ovarian data

Recap

We have a response of interest: continuous number, class type, survival time.

We want to know which features are associated with the response.

Note: we are implicitly assuming features are related through some functional form, e.g.

$$y = f(x; \beta) \tag{1}$$

Recap

For example: Linear regression

$$y = \sum_{j=1}^{p} \beta_j x_j \tag{2}$$

Logistic regression

$$\log \frac{p}{1-p} = \sum_{j=1}^{p} \beta_j x_j \tag{3}$$

Survival analysis (Cox's proportional hazards model)

$$h(t) = h_0(t) \exp \left[\sum_{j=1}^{p} \beta_j x_j \right] \tag{4}$$

Spike-and-Slab (SpSL) priors

Each coefficient β_j within our model has a corresponding latent variable z_j

z_j indicates whether the coefficient takes a value of 0 or not i.e. has an effect on our response

Formally,

$$\beta_j | z_j \overset{\text{ind}}{\sim} z_j \text{Laplace}(\lambda) + (1 - z_j)\text{Dirac}_0$$

$$z_j | w_j \overset{\text{ind}}{\sim} \text{Bernoulli}(w_j)$$

$$w_j \overset{\text{iid}}{\sim} \text{Beta}(a_0, b_0)$$

Updating our prior

We update our prior using Bayes theorem,

$$\underbrace{\Pi(\beta, z, w|\mathcal{D})}_{\text{Posterior}} \propto \underbrace{L(\mathcal{D}|\beta, z, w)}_{\text{Likelihood}} \times \underbrace{\Pi(\beta, z, w)}_{\text{SpSL prior}} \tag{5}$$

where $\mathcal{D}$ is our observed data, $\beta = (\beta_1, \ldots, \beta_p)^\top$, $z = (z_1, \ldots, z_p)^\top$ and $w = (w_1, \ldots, w_p)^\top$.

Giving a distribution for our coefficients constructed using observed data.

Posterior

The posterior $\Pi(\beta, w, z|\mathcal{D})$ is a rich mathematical object, giving us insight into the:

- Effect sizes β_j, given they are in the model, i.e. $\beta_j \neq 0$.
  - More specifically a distribution of effect sizes
- Inclusion probabilities w_j, the probability a feature is included in the model

Practical concerns

- Runtime
  - MCMC can take a very long time!
  - Addressed this via variational approximations
- Model selection criteria
- Prior parameters

Practical concerns

Runtime

- Maximum *a posterior* (MAP) estimates rather than the full posterior
- My work, we've pursued a different avenue using variational inference, which is slightly slower.

Model selection

- Standard approaches e.g. BIC
- Evidence lower bound (if using variational inference)

Practical concerns

Prior parameters

- If we do not have an *a priori* belief about the prior parameters tuning is important
- We can use the data to inform the choice e.g. maximizes the BIC or ELBO
- Methods can be sensitive to prior parameter choices e.g. BhGLM

An example on simulated data

We're going to be looking at simulated survival data where we have

- $n = 200$ (with 25% of times censored)
- $p = 1,000$ features
- $s = 10$ features that are non-zero

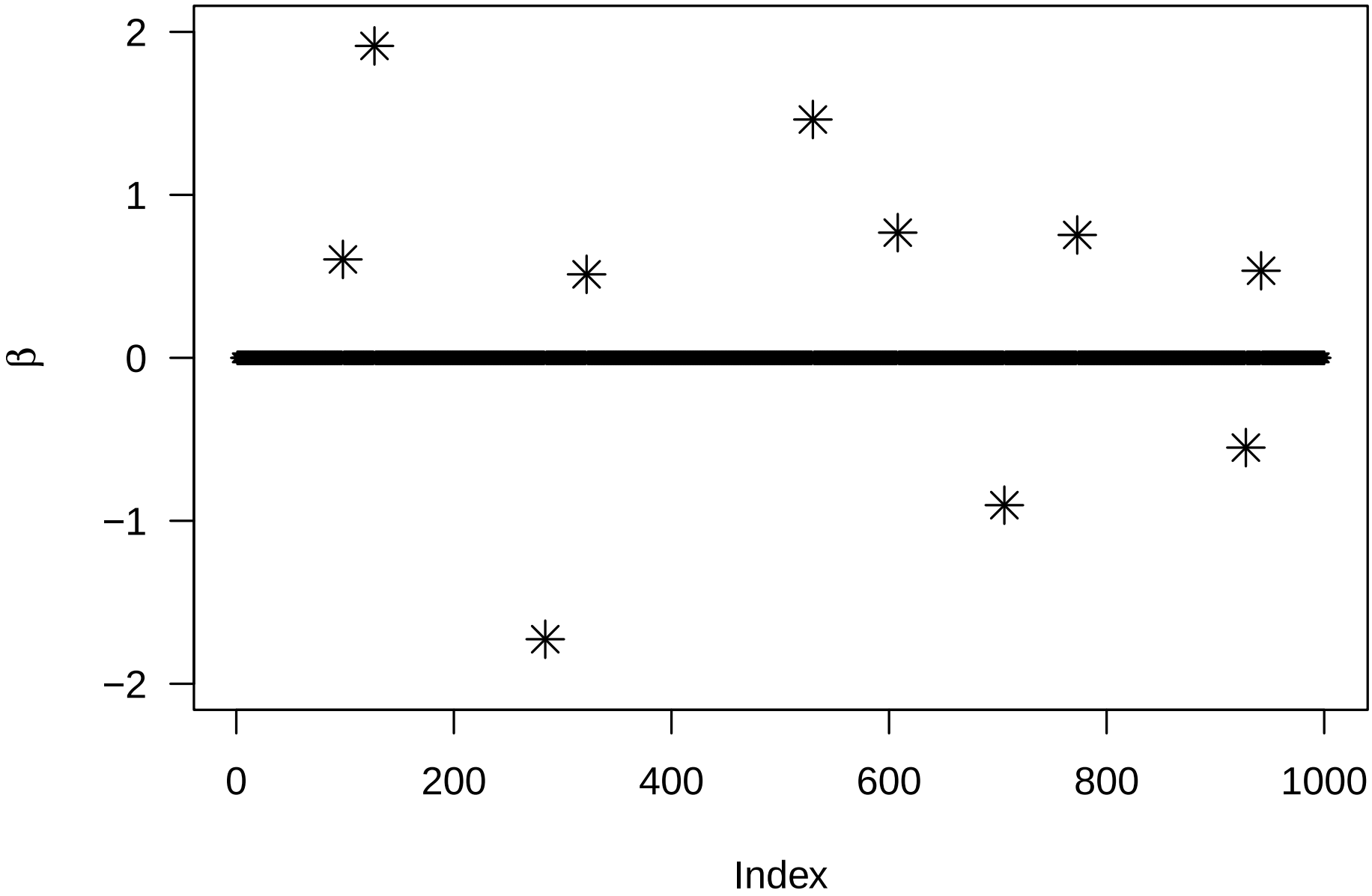
An example on simulated data

Specifically, our dataset is given as

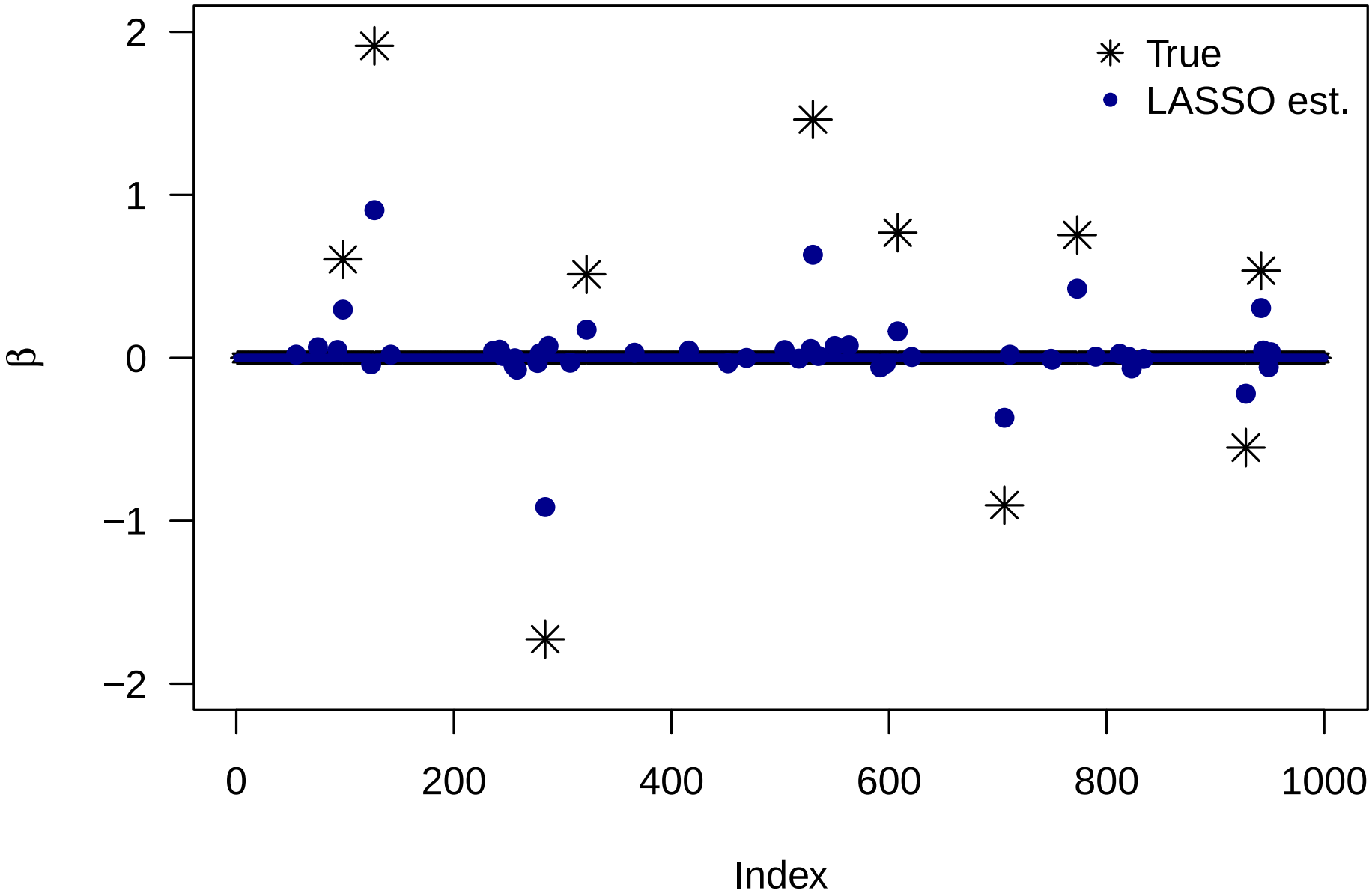
$$\mathcal{D} = \big(t_i, \quad \delta_i, \quad x_i\big), \quad \text{for } i = 1, \ldots, 200$$

where t_i is an observed time, δ_i is a censoring indicator, x_i are the predictors where $x_{i,j} \overset{iid}{\sim} N(0, 1)$

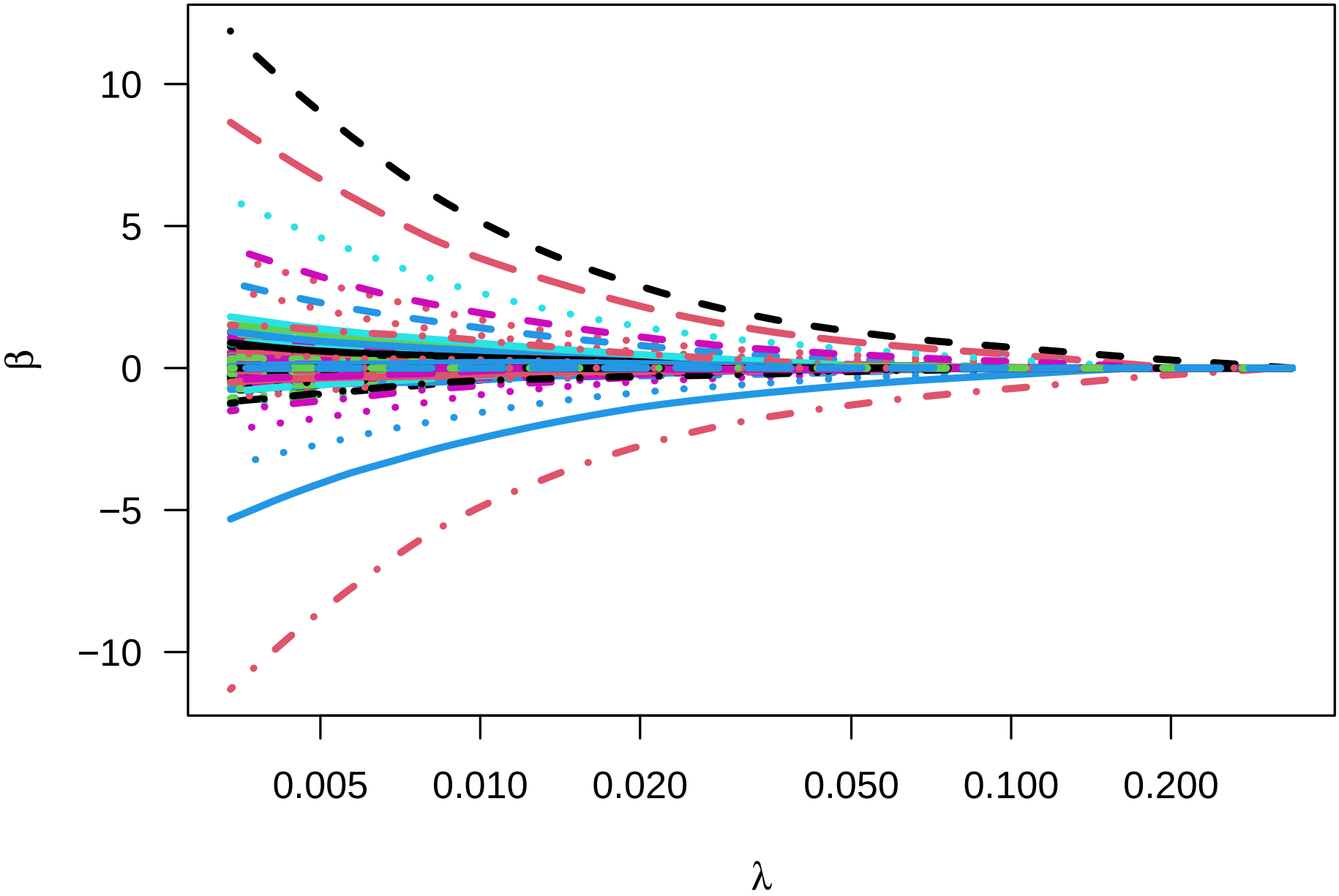
True effect sizes



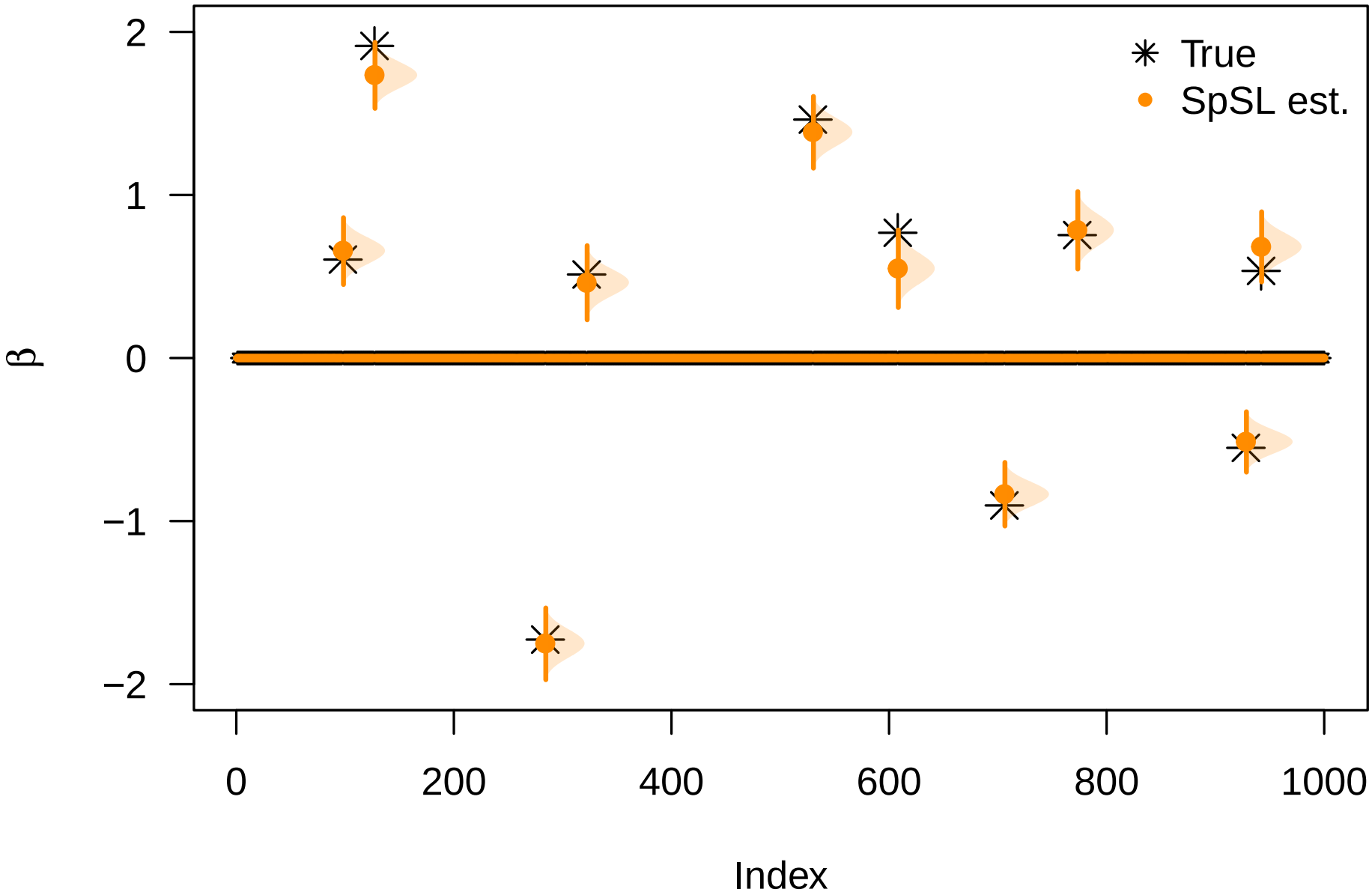
LASSO



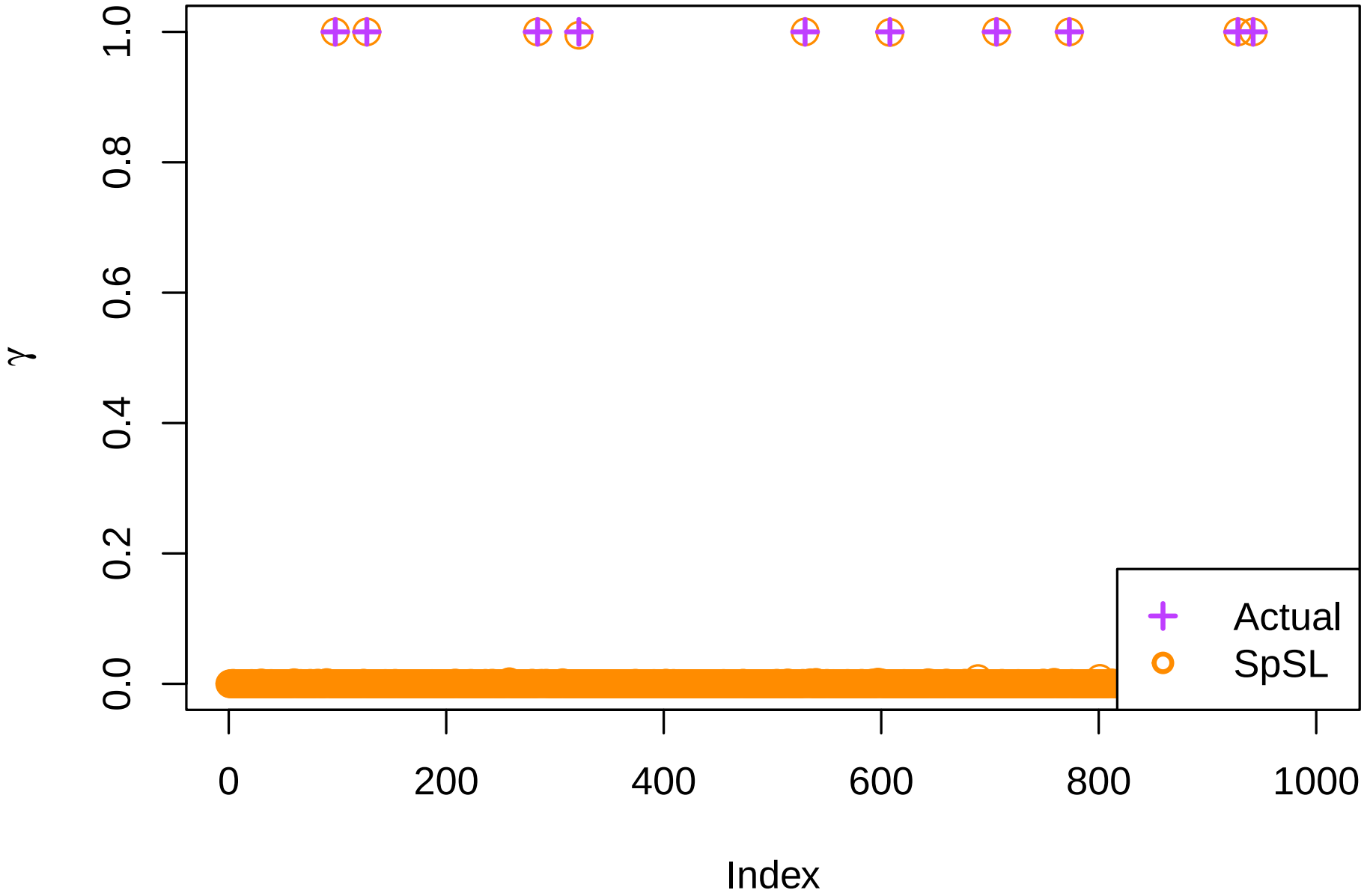
LASSO sensitivity



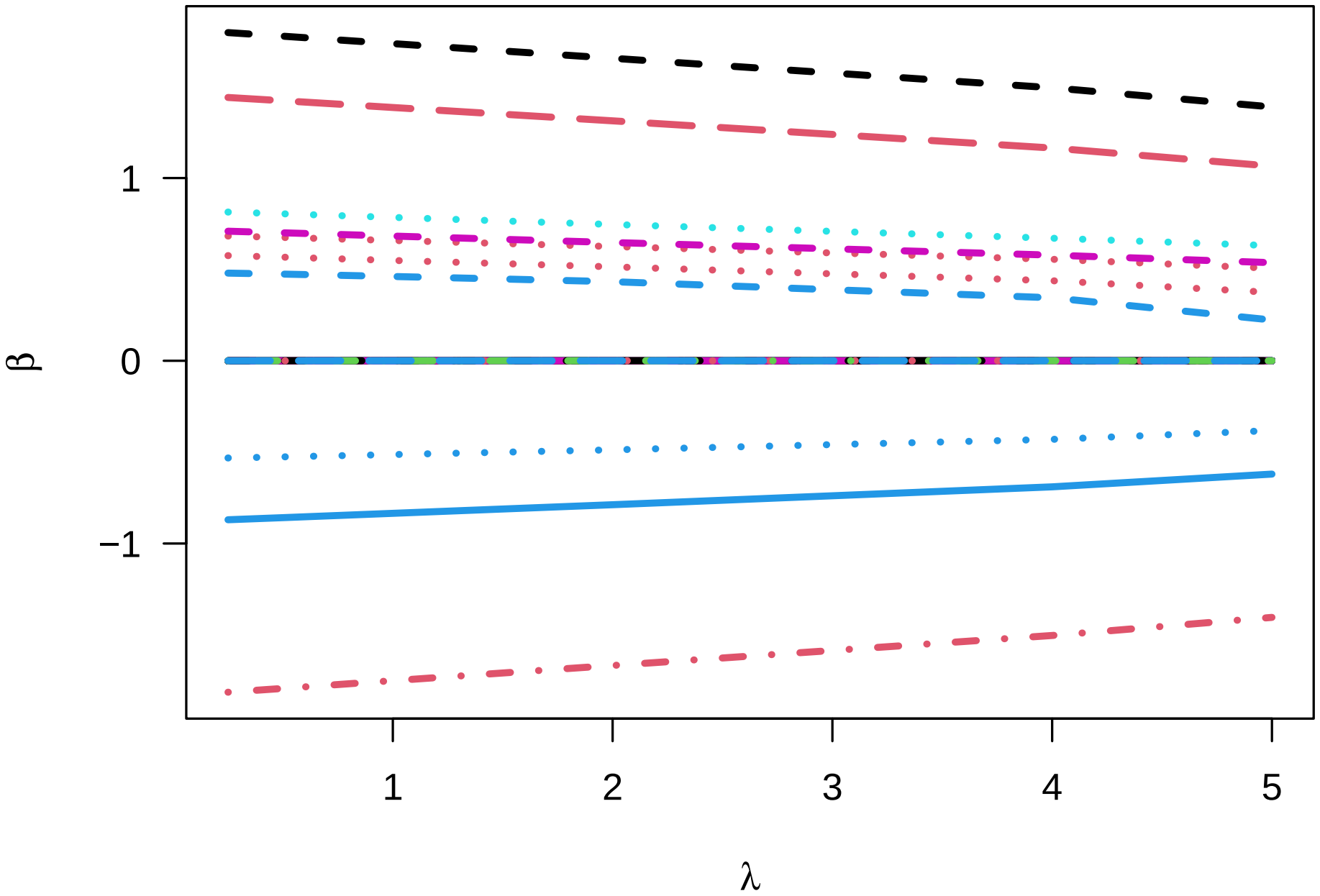
SpSL coefficient estimates



SpSL inclusion probabilities



SpSL sensitivity



SpSL v LASSO

	LASSO	SpSL
Good point estimates for β	×	✓
Uncertainty quantification	×	✓
Inclusion probabilities	×	✓
Not sensitive to regularization parameter	×	✓
Inclusion of *a prior* information	×	✓
Fast	✓	✓

A case study on real data

The dataset is a publicly available transcriptomic data where the response of interest is overall survival

- $n = 580$ with 39.5% censored
- $p = 12,042$

We're going to fit our variational approximation to the dataset.

Real data: cross validating

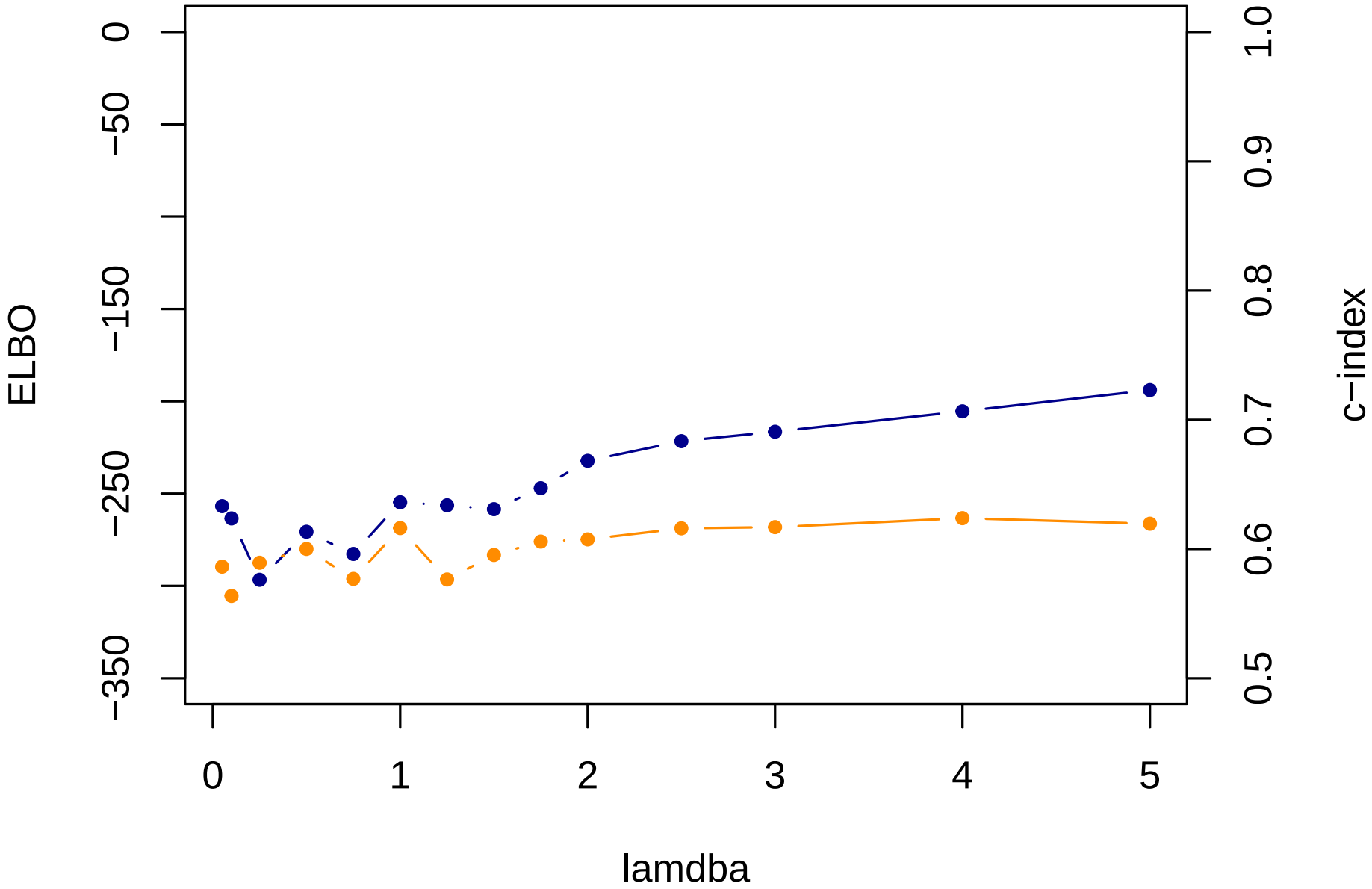
We have some *a priori* information about the number of non-zero coefficients so we set $a_0 = p/100$ and set $b_0 = p$

As we don't have *a priori information* about the coefficient magnitudes we cross validate to select the model parameter.

We use 10-fold cross validation and report the:

- Evidence lower bound
- C-index

Selecting the best model?



Variable selection

As our model isn't that sensitive to λ we use all the models to do variable selection.

We set the inclusion threshold to 0.5 and compute the proportion of times each feature is non-zero across the different models.

Variable selection

PI3	VSIG4	PPP3CA	IL7R	SDF2L1	D4S234E	DAP	CCR7
0.786	0.307	0.257	0.243	0.207	0.2	0.193	0.186
ACSL3	PLA2G2D	ADORA3	FLNA	SLAMF7	UBD	CD14	HABP2
0.157	0.157	0.121	0.121	0.107	0.107	0.086	0.086
LPXN	LCE2B	TBP	GALNT10	NOTCH4	RNF128	C5orf28	PPM2C
0.086	0.079	0.079	0.071	0.071	0.071	0.064	0.064
FJX1	TSPAN13	HSPB7	TREML2				
0.057	0.057	0.05	0.05				

Genes with biological interpretations discovered in the biomedical literature

(these are the only ones we checked!)

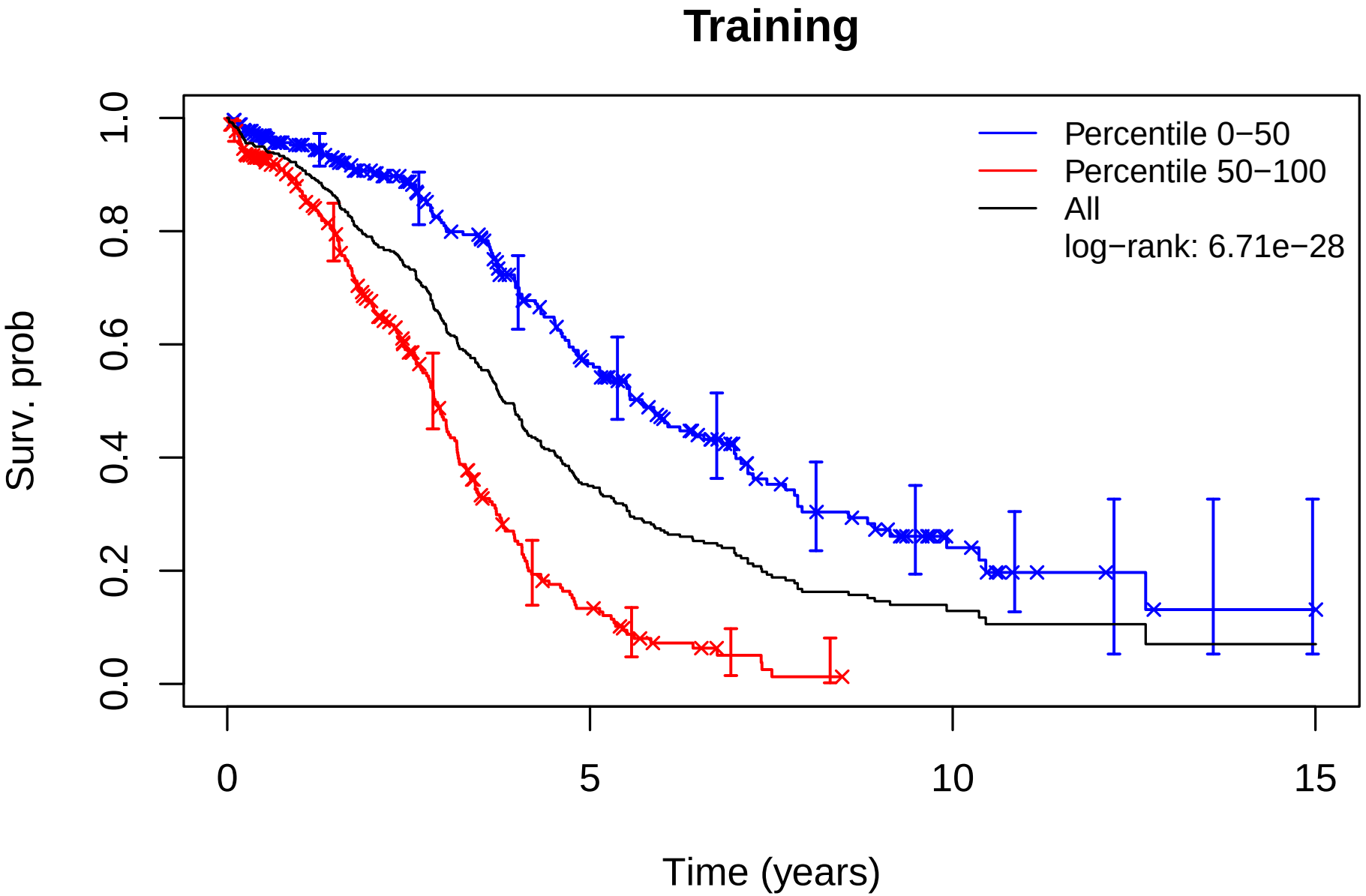
Prediction?

Prediction is fairly limited for survival models without further assumptions on the baseline hazard, $h_0(t)$.

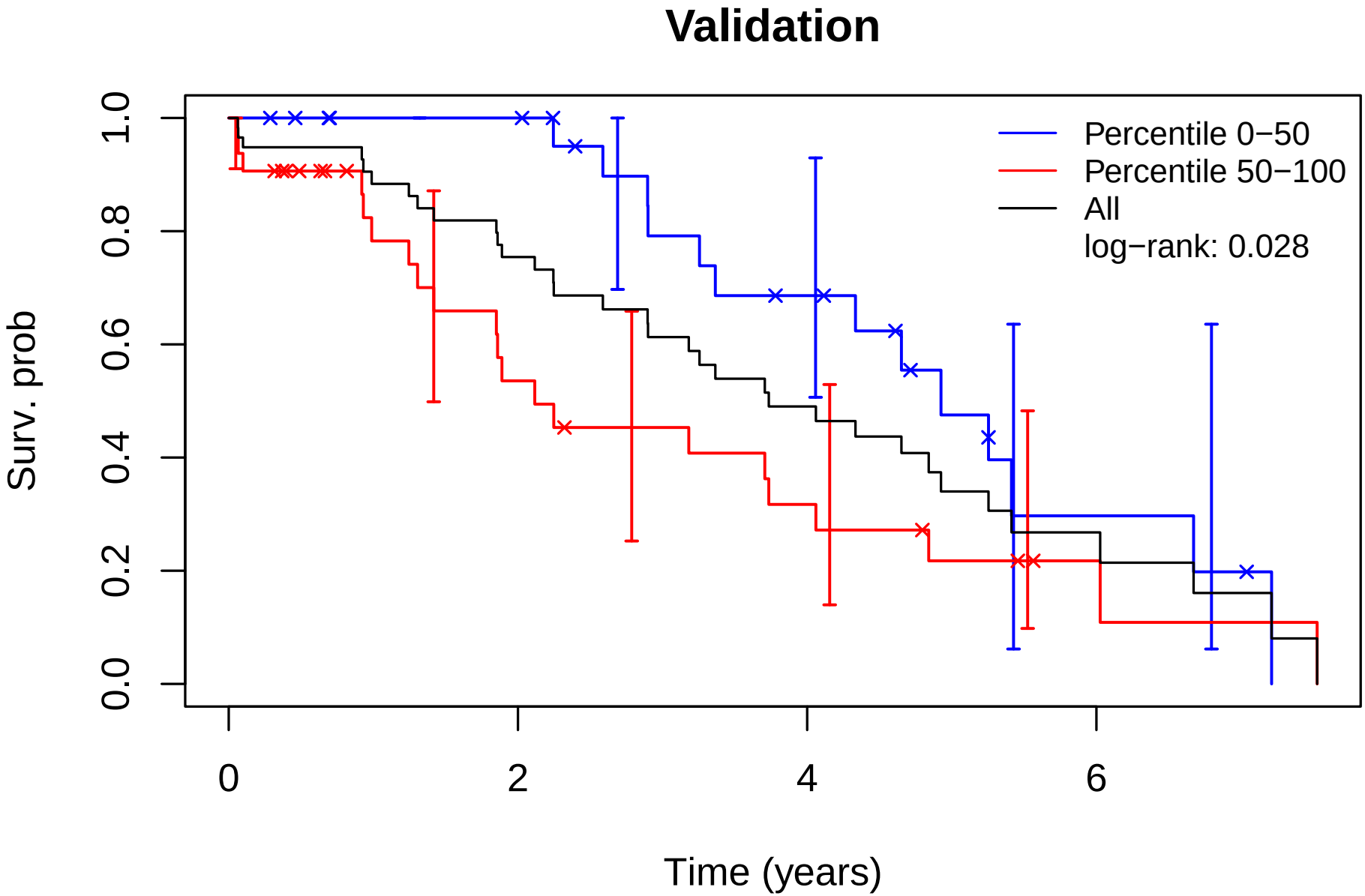
For survival models, the prognostic index is commonly used, given as

$$\eta_i = \sum_{j=1}^{p} \beta_j x_{ij} \tag{6}$$

Prediction?


Training
Surv. prob
0.0
0.2
0.4
0.6
0.8
1.0
0
5
10
15
Time (years)
Percentile 0−50
Percentile 50−100
All
log−rank: 6.71e−28

Prediction?


Validation
Surv. prob
0.0
0.2
0.4
0.6
0.8
1.0
0
2
4
6
Time (years)
Percentile 0−50
Percentile 50−100
All
log−rank: 0.028

R packages

R package available at

Variational Bayes for survival

`https://github.com/mkomod/survival.svb`

MCMC sampler

`https://github.com/mkomod/survival.ss`

Code for today's examples

`https://github.com/mkomod/presentations`

Reference

[Bis06] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006. Chap. 10.

[BKM17] David M. Blei, Alp Kucukelbir, and Jon D. McAuliffe. "Variational Inference: A Review for Statisticians". In: *Journal of the American Statistical Association* 112.518 (2017), pp. 859–877.

[GM93] Edward I. George and Robert E. McCulloch. "Variable Selection via Gibbs Sampling". In: *Journal of the American Statistical Association* 88.423 (1993), pp. 881–889.

[MB88] T. J. Mitchell and J. J. Beauchamp. "Bayesian variable selection in linear regression". In: *Journal of the American Statistical Association* 83.404 (1988), pp. 1023–1032.

Questions, Comments?

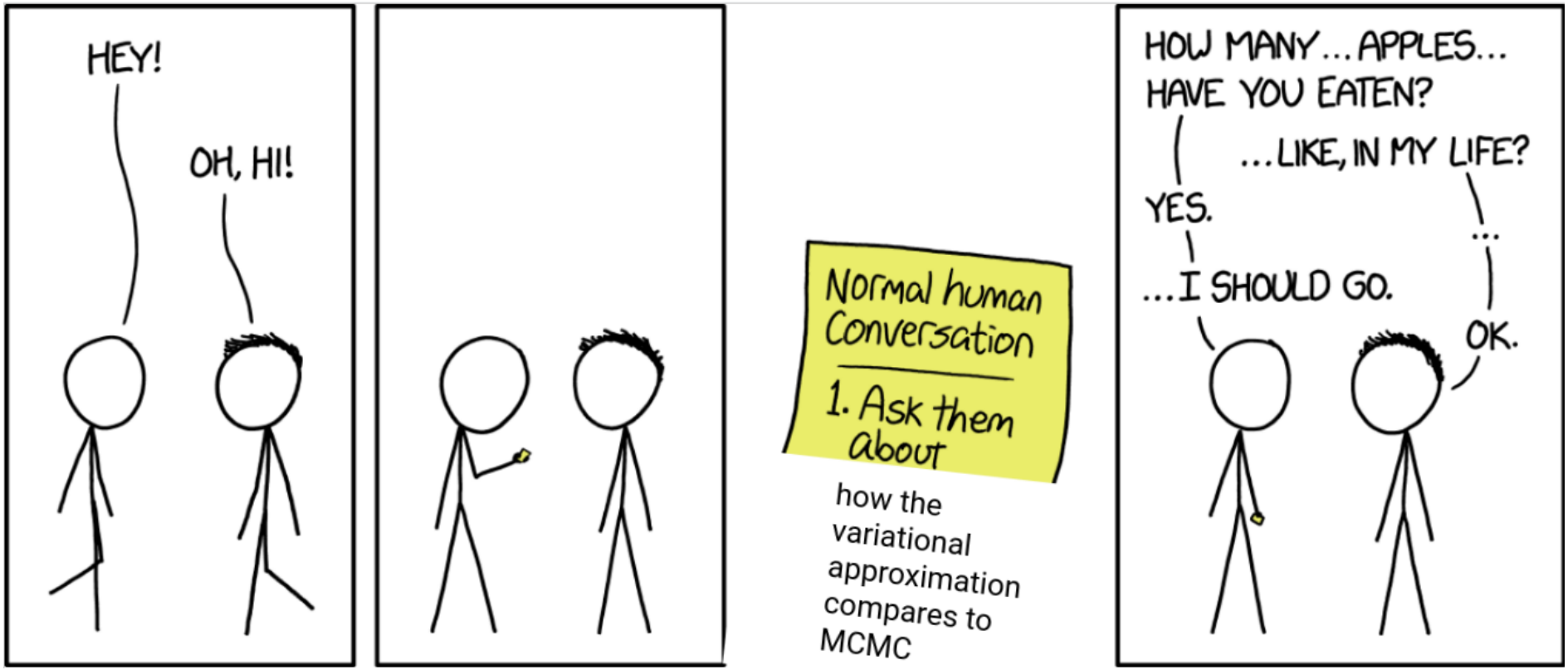


Figure: Source: `https://xkcd.com/1976/`, with some minor edits